PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Article submitted to journal

Keywords:
networks, epidemics, complex contagions

Author for correspondence:
Tim Rogers
e-mail: t.c.rogers@bath.ac.uk

Heterogeneous node responses to multi-type epidemics on networks

S. Moore[1] and T. Rogers[1]

[1]Centre for Networks and Collective Behaviour, Department of Mathematical Sciences, University of Bath, Bath, BA27AY, United Kingdom

Having knowledge of the contact network over which an infection is spreading opens the possibility of making individualised predictions for the likelihood of different nodes to become infected. When multiple infective strains attempt to spread simultaneously we may further ask which strain, or strains, are most likely to infect a particular node. In this article we investigate the heterogeneity in likely outcomes for different nodes in two models of multi-type epidemic spreading processes. For models allowing co-infection we derive message-passing equations whose solution captures how the likelihood of a given node receiving a particular infection depends on both the position of the node in the network and the interaction between the infection types. For models of competing epidemics in which co-infection is impossible, a more complicated analysis leads to the simpler result that node vulnerability factorises into a contribution from the network topology and a contribution from the infection parameters.

1. Introduction

The networks present everywhere in today's world are very busy. People travel rapidly across the globe carrying goods, ideas and diseases; inboxes are constantly bombarded with work and personal correspondence; social media facilitates the exchange of diverse ideas. Thus, when a contagion process of any kind breaks out it rarely acts in isolation; when ideas spread they may often inspire innovation, alternatives or opposing reactions. A curious example is given by trends in diet. In the United States in 2014 there was a sudden and massive increase, of up to 686% [1,2], in the consumption of a variety of alternative grains including Quinoa, Teff, Amaranth, Freekeh, Spelt and Kamut. This rapid change in consumer behaviour was driven in a large part

THE ROYAL SOCIETY PUBLISHING

by activity on social media networks. Crucially, there was a cooperative effect: individuals who had already received and propagated positive messages about one type of alternative grain are more likely to do the same for another, making the spread of excitement about these products inter-reinforcing. Contagious diseases can be similarly complex in their relationships. For example, the common cold may be attributed to some 200 different viruses and multiple bacterial infections most common amongst which, the Rhinovirus, is itself comprised of three different species and some 160 different types [3]. To understand a cold outbreak, one really needs study a host of different interacting, mutating and competing pathogens. Between-strain dynamics have a knock-on effect for control strategies. In pneumococcus, for example, the introduction of a multivalent vaccine has affected the dynamics of other strains [4].

Following its introduction by Kermark and McKendrick [5] in the context of modelling spread of infectious diseases in humans, the Susceptible–Infected–Recovered (SIR) epidemic model has been adapted to fit a host of different spreading dynamics, including viruses across computer networks, ideas through social media, and even traffic flow [6–8]. Furthermore, several multi-type infection models have also been considered with variants exhibiting cooperation [9], dependence [10,11] and competition [12].

The above cited works all consider the bulk behaviour of the propagating contagions. Looking in closer detail, knowledge of the underlying network structure over which a contagion spreads (see e.g. [13] for an introduction) should enable us to make individualised predictions for the likely outcomes for particular nodes [14]. Machinery to study heterogeneity in infection models has been in development for several years. An appreciation of the importance of network structure in disease spread, unseen by original mean field approaches [15], has inspired the emergence of network epidemiology [16–18]. One such approach is message passing, also known as the "cavity method", introduced in [19,20] for statistical physics models. The idea is to understand the vulnerability of, or conversely threat posed by, a single individual in the network by considering similar statistics of those in their immediate neighbourhood. This approach has been used, for example, by Newman and Karrer in [21] to describe the time development of an epidemic on a network, and in [14,22,23] to assess the heterogeneous responses of individual nodes.

In this article we adapt the message passing approach to study the heterogeneous node responses in a generalised multi-type infection model that may exhibit both cooperative characteristics and competitive ones. In particular, this enables us to determine a formula for node vulnerability in terms of the between-type infection probability matrix, T, and show that large outbreaks may only occur, in both competitive and non-competitive multi-type infection models, if the maximum eigenvalue of T is greater than the inverse of the largest eigenvalue of the network's Hashimoto matrix [24,25]. Furthermore, in the case of competitive infections we also show that node vulnerability to different infections factorises into a contribution from the network architecture, and a contribution from the infection dynamics (specifically, the dominant eigenvector of T).

We provide examples of our results implemented on both model networks (e.g. Erdős-Rényi random graphs) and those derived from human social behaviour (e.g. the *Epinions* social network). It is important to emphasise, however, that our work is not intended to directly apply to any particular network or contagion. This is a theoretical study which sheds new light on interesting possible behaviours of interacting epidemics, and the utility of the message passing approach in making detailed predictions of likely outcomes. As usual, certain assumptions (such as the fact that individuals differ only according to their position in the network, and not other factors) must be made in order to isolate the network-driven effects we are interested in exploring. Therefore, more detailed data-integrative modelling would required before our insights could be translated into policy that might seek to intervene in real-world interacting contagions.

The remainder of the article is organised as follows. In the next section we introduce a multi-type infection model with both competitive and non-competitive variants. Beginning with the most trivial single type instance of this model, we start by considering the dependency of edge

infection time on local topology, first when a neighbourhood's state is known deterministically and then in distribution in order to recap the formulation of the time dependent message passing equations as introduced previously by Karrer and Newman. We then move on to generalise this same approach to considering first non-competitive then competitive multi-type cases. Finally, a toy example of multi-type infection model where timings play a critical role to node vulnerabilities is considered. Here it is seen that careful choice of parameters leads to a change in type prevalence with infection speed; a phenomena however which is dependent upon graph sparsity.

2. Preliminaries

We present a general model for the spread of an infection with multiple types on a network, based on the well-known SIR process for simple infections, but with allowance for multiple different infection dynamics. We shall consider two cases: one allowing co-infection, where a single individual may be infected by more that one disease type, and another in which infection with one type excludes infection with any other. Our model is chosen to be as simple as possible, while retaining the key features (network topology and multi-type infections) that we wish to investigate. Many complicating features would need to be incorporated if future research on models of this type is to make actionable predictions about a given real-world contagion, but this is beyond the scope of the present theoretical investigation.

Our simple disease model is constructed as follows. We work with a static population of individuals with potential infectious contacts described by some fixed network. With respect to each of some number of different disease strains (or *types*), every individual node is in one of three possible states: susceptible, infected, or recovered. The model evolves in time with each individual infected with a strain α either recovering from that strain with rate r^α, or transmitting an infection to a neighbour. Importantly, rather than the separate types being treatable as separate infections, we take a more general approach in which each strain may in principle pass on any other strain. An individual carrying infection type α will attempt to induce an infection of type β in a neighbour with rate $f^{\beta\alpha}$. Whether or not this infection attempt succeeds is dependent on the target's own infection history. We consider two cases: co-infection, and competitive exclusion.

If co-infection is allowed, each individual node may contract multiple strains of the epidemic, potentially from multiple infected neighbours. In the exclusive case an individual may only be able to contract a maximum of one infection strain which adds an element of competition to the model, making it possible for a virulent infection to crowd-out a weaker strain. This creates the complication that in order to determine whether an individual is infected by a particular strain, we the need to know not just whether an individual is reached by that infection type, but also which type reaches them first. This makes it crucial to consider the full time development of the multi-type epidemic in order to determine the end state.

(a) Message-passing

To analyse the node-dependent outcomes of the multi-type epidemic model introduced above, we will employ a variant of the message-passing approach formulated by Newman and Karer [21]. Let us begin by recapping here the derivation of the method for the case of a single infection type.

For a single-type infection we may consider the time $t_{i\leftarrow j}$ at which node j attempts to pass the infection to i, a quantity which is infinite if either j never receives the infection themselves or recovers before attempting to pass it on. This quantity is equal to the first time at which one of j's other neighbours attempts to pass it the infection, plus an additional time period, S_{ij}, before j will itself transmit to i (note $S_{ij} = \infty$ if j recovers before it can propagate the infection). That is

$$t_{i\leftarrow j} = \min_{k\in\mathcal{N}(j)\setminus i} \left[t_{j\leftarrow k}\right] + S_{ij}\,. \tag{2.1}$$

The time until infection of node i can then be deduced by computing the first time at which one of i's neighbours attempts to pass on the infection to it, that is,

$$t_i = \min_{j \in \mathcal{N}(i)} [t_{i \leftarrow j}] \,. \tag{2.2}$$

From these definitions we can formulate equations for the distribution of $t_{i \leftarrow j}$. More specifically, we shall consider the probability, $H_{i \leftarrow j}(t)$, that i does not receive the infection from j before time t. In order to make progress here we shall introduce a sparse tree-like network approximation. If the graph is tree-like then the probable infection times of i's neighbours are approximately independent; using equation (2.2) we compute that the probability i doesn't receive the infection before time t is given by

$$H_i(t) = \prod_{j \in \mathcal{N}(i)} H_{i \leftarrow j}(t) \,. \tag{2.3}$$

To find an expression for $H_{i \leftarrow j}(t)$ we must integrate over the possible values of the delay S_{ij}. Write p for the density function of the i.i.d random variables S_{ij}. In a single type infection with recovery rate r and transmission rate f, we have $p(s) = fe^{-(r+f)s}$. Equation (2.1) then gives

$$H_{i \leftarrow j}(t) = 1 - \int_0^t \left(1 - \prod_{l \in \mathcal{N}(j) \setminus i} H_{j \leftarrow l}(t-s) \right) p(s) \, \mathrm{d}s \,. \tag{2.4}$$

This expression is substantially simplified in the long-time limit. Taking $t \to \infty$, we may use this system to assess how likely each node is to ever be reached by the infection. To this end, write $T = \int_0^\infty p(s) \mathrm{d}s = f/(f+r)$ to be the probability that an infected individual will ever attempt to pass on the infection to a given neighbour. Then we may consider the probability, $h_{i \leftarrow j} = H_{i \leftarrow j}(\infty)$, that j will ever attempt to infect i by taking $t \to \infty$ in (2.4) to give

$$h_{i \leftarrow j} = 1 - T \left(1 - \prod_{k \in \mathcal{N}(j) \setminus i} h_{j \leftarrow k} \right) \,. \tag{2.5}$$

We define the *vulnerability* of node i, v_i, to be the probability it becomes infected eventually. From equations (2.3) and (2.5), we deduce:

$$v_i = 1 - \prod_{j \in \mathcal{N}(i)} h_{i \leftarrow j} \,. \tag{2.6}$$

To summarise, by first formulating an expression for t_{ij} given complete knowledge of the time progression of j's neighbourhood, then considering this in probability to find $H_{ij}(t)$ and finally sending t to infinity to find h_{ij} we are able to calculate h_i, the probability that i will never be reached by an infection. This is a powerful method for analysing the node response to epidemics on networks, the question we now wish to address is how it might be applied to more complex models.

3. Co-infection

For a co-infecting multi-type model, where a single individual may be able to gain any number of types, we seek to investigate the heterogeneous node responses for each type, as described by the probability v_i^α that individual i becomes infected by some specific strain α at any point during the outbreak. We refer to this as the vulnerability of node i to strain α.

Our analysis proceeds in a generalisation of the message passing approach. The time at which a node j will attempt to send the α strain to i will be equal to the minimum times at which any of j's other neighbours pass on any strain β to j, plus the delay $S^{\beta\alpha}$ between j acquiring strain β

and passing on strain α. This logic encodes the statement:

$$t^{\alpha}_{i\leftarrow j} = \min_{\beta}\left[\min_{k\in\mathcal{N}(j)\setminus i}\left[t^{\beta}_{j\leftarrow k}\right] + S^{\beta\alpha}_{i\leftarrow j}\right]. \tag{3.1}$$

Note that $t^{\alpha}_{i\leftarrow j} = \infty$ if no strain that succeeded in infecting j then goes on to pass strain α to i. It is important to remark here that in writing (3.1) we make the assumption that infection of node i with a certain strain does not affect the transmission of other types. In the next section we will consider the effect of complete competitive exclusion (i.e. each node may be infected with at most one strain); the intermediate case of partial cross-immunity is far more complex.

An approximate form of the complementary cumulative distribution function $H^{\alpha}_{i\leftarrow j}(t)$ of $t^{\alpha}_{i\leftarrow j}$ may be obtained from (3.1) in a straightforward generalisation of (2.4). Specifically,

$$H^{\alpha}_{i\leftarrow j}(t) \approx 1 - \int_0^t \sum_{\beta}\left(1 - \prod_{l\in\mathcal{N}(j)\setminus i} H^{\beta}_{j\leftarrow l}(t-s)\right) p^{\beta\alpha}(s)\,\mathrm{d}s\,, \tag{3.2}$$

where $p^{\beta\alpha}$ is the density function of $S^{\beta\alpha}$. In writing (3.2) we make the assumption that contributions to the infection pressure at a given node from strains of different types are approximately additive. This holds in various circumstances including when t is small, when the outbreak is near critical, when mutations between types are relatively rare, or when one type spreads faster than the others. In the long term, the probability $h^{\alpha}_{i\leftarrow j}$ that i will never be infected with strain α by j is given by:

$$h^{\alpha}_{i\leftarrow j} = 1 - \sum_{\beta} T^{\beta\alpha}\left(1 - \prod_{k\in\mathcal{N}(j)\setminus i} h^{\beta}_{j\leftarrow k}\right), \tag{3.3}$$

where

$$T^{\beta\alpha} = \int_0^{\infty} p^{\beta\alpha}(t)\,\mathrm{d}t = \frac{f^{\beta\alpha}}{r^{\beta} + \sum_{\gamma} f^{\beta\gamma}}\,.$$

This quantity defines the chance of a type β individual ever attempting to send a infection α to a neighbour. We will refer to the matrix of all these values for different α, β as the *transmission matrix*. Finally, the vulnerability v^{α}_i of node i to strain α will be given by

$$v^{\alpha}_i = 1 - \prod_{j\in\mathcal{N}(i)} h^{\alpha}_{i\leftarrow j}\,. \tag{3.4}$$

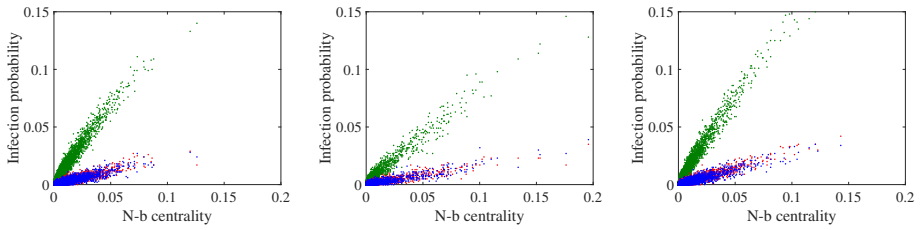
To numerically compute node vulnerabilities, it is first necessary to solve the system (3.3). In doing this it is convenient to examine the network of relationships between variables $h^{\alpha}_{i\leftarrow j}$. The non-backtracking, or Hashimoto, graph of directed edges is defined by the relation that $e = (i \to j)$ links to $e' = (k \to \ell)$ if $k = j$ but $\ell \neq i$ [26]. On this graph the system (3.3) then equivalently becomes

$$h^{\alpha}_{e} = 1 - \sum_{\beta} T^{\beta\alpha}\left(1 - \prod_{e'\in\mathcal{N}(e)} h^{\beta}_{e'}\right), \tag{3.5}$$

where $\mathcal{N}(e)$ refers to neighbours of edges in the Hashimoto graph. Let us write B for the adjacency matrix of the Hashimoto graph, known as the *non-backtracking matrix*.

Solutions to the system of equations (3.5) may be found by iteration to give the probabilities over directed edges which can be used to find the node vulnerabilities, h^{α}_i, by equation (3.4). There is a trivial solution $h^{\alpha}_e \equiv 1$, corresponding to all nodes never receiving any infection. Depending on the values of $T^{\beta\alpha}$, there may or may not be a non-trivial solution set.

The boundary of the region of parameter space in which a non-trivial solution exists (and hence where a large-scale outbreak is possible) is known as the percolation threshold, which we now compute. We proceed by analysing the linear stability of the no-infection state $\mathbf{h} = 1$. To this



(a) (b) (c)

Figure 1: Plots showing estimated infection probabilities for three coevolving non-exclusive infection strains against non-backtracking centrality for nodes on a variety of real networks. In each plot, the individual node infection probabilities are taken as the mean number of infections of each type in 1000 simulated runs from random starting nodes. The infection dynamics match those used in equations (3.10)-(3.12) tuned to be just slightly above the percolation threshold, with strain 1 shown in blue, strain 2 in red and strain 3 in green. The networks used are (a) a trust network of 75879 nodes from the *Epinions* social network [28] (b) a friendship network from the music streaming service *Deezer* [29] (c) A network of internal email communication within the company *Enron* [30].

end taking derivatives from (3.5) gives us

$$\frac{\partial h_e^\alpha}{\partial h_{e'}^\beta} = \mathbf{1}_{e' \in \mathcal{N}(e)} T^{\beta\alpha} \prod_{e'' \in \mathcal{N}(e) \setminus e'} h_{e''}^\beta \,, \tag{3.6}$$

and therefore the Jacobian matrix of the message-passing system at the trivial state is given by

$$J_{e,e'}^{\beta,\alpha} = \left. \frac{\partial h_e^\alpha}{\partial h_{e'}^\beta} \right|_{\mathbf{h}=1} = B_{e,e'} T^{\beta\alpha} \,. \tag{3.7}$$

If J has an eigenvalue with real part larger than one, then the dynamical system described by iteration of the message-passing equations is unstable around the state $\mathbf{h} = 1$, and an outbreak is possible. According to (3.7), J is the Kronecker product of the matrices B and T. Both of these have non-negative entries and hence real maximum eigenvalues, say λ_B and λ_T, respectively. We deduce that a multi-type infection defined by transmission matrix T spreading on a network with non-backtracking matrix B has a non-zero probability to result in a large outbreak if and only if

$$\lambda_T \lambda_B > 1. \tag{3.8}$$

Around the neighbourhood of percolation, the factorisation we find in equation (3.7) implies that the node vulnerabilities will be proportional to their *non-backtracking centrality* [27], as expounded in [22]. Moreover, it may be seen that by augmenting this method to include specific type vulnerabilities, by way of the matrix T, it may be found that the probability node i becomes infected by strain α will be proportional to

$$v_i^\alpha \sim c_i w^\alpha \,, \tag{3.9}$$

where w is the top eigenvector of T and c_i the non-backtracking centrality of node i.

This proportionality is demonstrated in Figure 1 on a variety of real world networks for an infection with three types. For the case that λ_T is significantly larger than λ_B the relationship is more complex, and full solution of the system (3.5) is necessary.

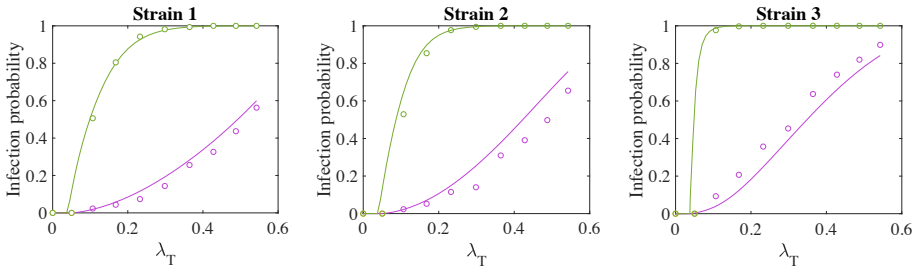


Figure 2: Vulnerabilities to an infection with three types, computed for two separate nodes in a real world network of nodes corresponding to 14113 company Facebook pages with edges representing mutual likes among them [29]. Solid lines show the theoretical prediction of equations (3.5) and (3.4); circles show the result of simulations averaged over 1000 samples. The chosen nodes have degree 3 (magenta, lower) and degree 40 (green, higher), and show a strong differential in outcome.

Example: three types

Let us consider a co-infection model with infection dynamics of the form

$$f^{\beta\alpha} = b^{\beta} m^{\beta\alpha}\,, \tag{3.10}$$

where b^{β} decribes the overall transmission rate for type β, and $m^{\beta\alpha}$ the probability of mutation from type β to type α during transmission (with the rule $\sum_{\gamma} m^{\beta\gamma} = 1$). For simplicity we fix all recovery rates as equal, so $r^{\beta} \equiv r$. This gives the transmission matrix

$$T^{\beta\alpha} = \frac{b^{\beta}}{r + b^{\beta}} m^{\beta\alpha}\,. \tag{3.11}$$

Figure, 2 shows example results of such an infection. Here we have used the mutation matrix

$$m = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}\,, \tag{3.12}$$

and infection rates $b^1 = 0.1x$, $b^2 = 0.2x$, $b^3 = x$, where x is a parameter we vary in order to achieve different values of λ_T for plotting.

4. Exclusive infections

Many contagion processes spread at the expense of excluding others. In our model, if we do not permit co-infection, then the time at which j attempts to send the α strain to i depends strongly on which strain has suceeded in infecting j. Specifically,

$$t^{\alpha}_{i\leftarrow j} = \min_{\gamma,k}\left[t^{\gamma}_{j\leftarrow k}\right] + S_{\alpha\beta^{\star}} \tag{4.1}$$

where $\beta^{\star} = \arg\min_{\gamma}\left[\min_{k}\left[t^{\gamma}_{j\leftarrow k}\right]\right]$ is the strain infecting j. From here, to calculate $H^{\alpha}_{i\leftarrow j}(t)$ given the equivalent distributions for neighbouring nodes as before, we take the approach of first calculating how likely any infection is to be transmitted down the edge in question before then assessing the strain transmitted conditioned on this value.

Take now $q^{\alpha}_{i\leftarrow j}(t)$ to be the density function for the distribution describing the likelihood of j passing on α to i, so

$$H^{\alpha}_{i\leftarrow j}(t) = 1 - \int_0^t q^{\alpha}_{i\leftarrow j}(s)\mathrm{d}s\,. \tag{4.2}$$

Now if we consider the equivalent distribution for j passing on anything to i, with probability density function $q_{i\leftarrow j}(t)$, since types are exclusive it follows that

$$q_{i\leftarrow j} = \sum_{\alpha} q^{\alpha}_{i\leftarrow j} \tag{4.3}$$

and further there exists weights $\omega^{\alpha}_{i\leftarrow j}$ so that

$$q^{\alpha}_{i\leftarrow j} = \omega^{\alpha}_{i\leftarrow j} q_{i\leftarrow j}\,. \tag{4.4}$$

Thus

$$H^{\alpha}_{i\leftarrow j}(t) = 1 - \omega^{\alpha}_{i\leftarrow j}\int_0^t q_{i\leftarrow j}(s)\mathrm{d}s \tag{4.5}$$

$$= 1 - \omega^{\alpha}_{i\leftarrow j}\left(1 - H_{i\leftarrow j}(t)\right) \tag{4.6}$$

Now for i to pass on α to j at time t, some other neighbour of j, say k, must pass on an infection of some given type, say β, at some prior time $t-\tau$ and then we must have a successful α infection from this after the remaining time period τ, this being described by the density $p^{\alpha\beta}$. To additionally observe exclusivity j must not gain an infection via another node before k, this happens with probability $\prod_{l\in\mathcal{N}(j)\setminus k} H_{i\leftarrow j}(t-\tau)$. Combining all this we see that

$$q^{\alpha}_{i\leftarrow j}(t) = \int_0^t \sum_{\substack{\alpha\\ k\in\mathcal{N}(j)}} q^{\beta}_{j\leftarrow k}(t-\tau) \prod_{l\in\mathcal{N}(j)\setminus k} H_{i\leftarrow j}(t-\tau)p^{\alpha\beta}(\tau)\,\mathrm{d}\tau \tag{4.7}$$

and further by (4.2) it follows that

$$H^{\alpha}_{i\leftarrow j}(t) = 1 - \int_0^t\int_0^s \sum_{\substack{\alpha\\ k\in\mathcal{N}(j)}} q^{\beta}_{j\leftarrow k}(s-\tau) \prod_{l\in\mathcal{N}(j)\setminus k} H_{i\leftarrow j}(s-\tau)p_{\alpha\beta}(\tau)\,\mathrm{d}\tau\mathrm{d}s\,. \tag{4.8}$$

Then by (4.4) and substituting $u = s-\tau$

$$H^{\alpha}_{i\leftarrow j}(t) = 1 - \int_0^t \sum_{\substack{\alpha\\ k\in\mathcal{N}(j)}} p^{\alpha\beta}(\tau)\omega^{\beta}_{j\leftarrow k}\int_0^{t-\tau} q_{j\leftarrow k}(u) \prod_{l\in\mathcal{N}(j)\setminus k} H_{i\leftarrow j}(u)\,\mathrm{d}u\mathrm{d}\tau\,. \tag{4.9}$$

Now by summing over all the neighbours of j the inner integral here gives the probability that j gains any infection from any neighbour before time $t-\tau$, that is

$$\sum_{k\in\mathcal{N}(j)}\int_0^{t-\tau} q_{j\leftarrow k}(u) \prod_{l\in\mathcal{N}(j)\setminus k} H_{i\leftarrow j}(u)\,\mathrm{d}u = 1 - \prod_{k\in\mathcal{N}(j)} H_{j\leftarrow k}(t-\tau) \tag{4.10}$$

and thus

$$H^{\alpha}_{i\leftarrow j}(t) = 1 - \int_0^t \sum_{\beta} p^{\alpha\beta}(\tau)\omega^{\beta}\left(1 - \prod_{k\in\mathcal{N}(j)\setminus i} H_{j\leftarrow k}(t-\tau)\right)\mathrm{d}\tau\,. \tag{4.11}$$

We now have a similar expression to (4.1), the equivalent formulation for non-exclusive types, the difference here being due to the fact that the only infection that i may be infected by is the first to reach j.

In this way we see the perhaps surprising result that the probability of an individual gaining a specific type of infection may be broken down into two independent parts describing the probability of gaining any type of infection together with the probability that the infection if gained is of a certain type.

This decomposition is useful in computing probability $h^{\alpha}_{i\leftarrow j}$ that i will never get infected by strain α from j. The probability that j will send an α message to i having itself received a β

message is given by

$$T_{\alpha\beta}\omega^{\beta}\left(1-\prod_{k\in\mathcal{N}(j)}h_{j\leftarrow k}\right) \tag{4.12}$$

and so the probability of i ever receiving an α infection from j will be given by

$$\omega^{\alpha}(1-h_{i\leftarrow j})=\sum_{\beta}T_{\alpha\beta}\omega^{\beta}\left(1-\prod_{k\in\mathcal{N}(j)\setminus i}h_{j\leftarrow k}\right). \tag{4.13}$$

Writing $\boldsymbol{\omega}$ for the vector with entries ω_α we can deduce the eigenvector equation

$$T\boldsymbol{\omega}=\lambda_T\boldsymbol{\omega}\,, \tag{4.14}$$

where it must hold that

$$\lambda_T=\frac{1-h_{i\leftarrow j}}{1-\prod_{k\in\mathcal{N}(j)\setminus i}h_{j\leftarrow k}}\,. \tag{4.15}$$

That is, $\boldsymbol{\omega}$ is an eigenvector of the matrix T with eigenvector λ_T, and we may rearrange (4.15) to obtain the message-passing equations

$$h_{i\leftarrow j}=1-\lambda_T\left(1-\prod_{k\in\mathcal{N}(j)\setminus i}h_{j\leftarrow k}\right), \tag{4.16}$$

with the rule

$$h^{\alpha}_{i\leftarrow j}=\omega^{\alpha}h_{i\leftarrow j}\,. \tag{4.17}$$

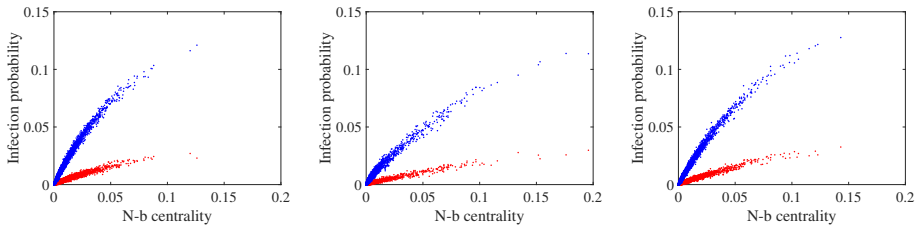
Note that the system (4.16) is precisely the same as the well studied case of equation (2.5), with the substitution $T\mapsto\lambda_T$.

We have thus mapped the analysis of the initially more complex exclusive infection model onto the previous case, allowing one to immediately translate existing results. For example, we find that the percolation threshold is once again given by the condition $\lambda_T\lambda_B>1$; the same as for the co-infection case. This result is perhaps to be expected as at the limit of infection survival each individual will become exposed to at most one infection. This also means that again, around the neighbourhood of percolation, the node vulnerabilities will be proportional to their non-backtracking centrality, and equation (3.9) holds similarly to the non-exclusive case– this relationship demonstrated in Figure 3, with plots comparable to Figure 1.

Example: speed vs reliability

In a contagion with competing strains it may be interesting to know what characteristics will become dominant. Here we consider a trade-off between the speed and reliability with which an infection may spread. Both these quantities may be expected to have a bearing on whether a particular strain becomes dominant in a population. Using our formulated message passing equations we can then seek to answer the question of what makes for a more successful contagion type; that which is slow but reliable or fast but unreliable

Consider a model with two competing strains, α and β, with dynamics specified by infection and mutation rates as presented in equations (3.10)-(3.12). The case we are interested in is when $b^{\alpha}<b^{\beta}$, but $m^{\beta\alpha}>m^{\alpha\beta}$; that is, β spreads faster, but is also more likely to mutate. In the two-type



(a) (b) (c)

Figure 3: Plots showing estimated infection probabilities for two coevolving non-exclusive infection strains against non-backtracking centrality for nodes on a variety of real networks. In each plot, the individual node infection probabilities are taken as the mean number of infections of each type in 5000 simulated runs from random starting nodes. The infection dynamics are chosen to be just above the percolation threshold, with strain α shown in blue, and strain β in red, mutation probabilities $m^{\alpha\beta} = 0.1$, $m^{\beta\alpha} = 0.25$ and relative speed $b^\beta/b^\alpha = 4$. The networks used are (a) a trust network of 75879 nodes from the *Epinions* social network [28] (b) a friendship network from the music streaming service *Deezer* [29] (c) a network of internal email communication within the company *Enron* [30].

case, the matrix T has the form

$$T = \begin{pmatrix} \frac{m^{\alpha\alpha}}{1+d/b^\alpha} & \frac{m^{\alpha\beta}}{1+d/b^\alpha} \\ \frac{m^{\beta\alpha}}{1+d/b^\beta} & \frac{m^{\beta\beta}}{1+d/b^\beta} \end{pmatrix} \tag{4.18}$$

The top eigenpair $(\lambda_T, \boldsymbol{\omega})$ of T can be computed directly, then respectively used in equation (4.16) to compute node vulnerabilities and the overall probability of an outbreak, and in equation (4.17) to determine which infection type is dominant.

In Figure. 4 we plot the fraction of infections achieved by each strain in an Erdős-Rényi random graph. The plot shows how this quantity varies as a function of relative speed of infections b^β/b^α, with mutation probabilities are fixed at $m^{\alpha\beta} = 0.01$ and $m^{\beta\alpha} = 0.25$, and death rate $d = 2$. As is clearly visible from the figure, an infection which reproduces less reliably can come to dominate a more reliable one, provided it spreads fast enough. This relationship is considered in more detail in Figure. 5, shows a phase diagram of outcomes (no outbreak, α dominates, β dominates) as both strain infection rates are varied. In that figure we use $m^{\alpha\beta} = 0.1$ and $m^{\beta\alpha} = 0.33$.

5. Conclusion

In this article we have constructed systems of equations capturing the time development of node states during a complex multi-type infection outbreak, dependent upon the local network topology. Furthermore, we have shown that this is an effective method for ranking nodes by their eventual susceptibility to a complex infection outbreak. This makes for a framework to be applied to evaluate heterogeneity of individual node outcomes in many complex contagion models.

Application of our co-infection model is, for instance, of potential interest in understanding the dynamics of cooperative contagions. These have been studied for example in [9] to model phenomena such as the co-epidemic of pneumonia and the 1918 Spanish flu. Alternatively, on might apply our approach to help understand individual strategy in models for human, Web based, cooperation as studied in [31]. Meanwhile, the framework for exclusive (i.e. competitive) contagions may be applied to study individual strain vulnerability in infection models with cross immunity, as studied in [32,33].

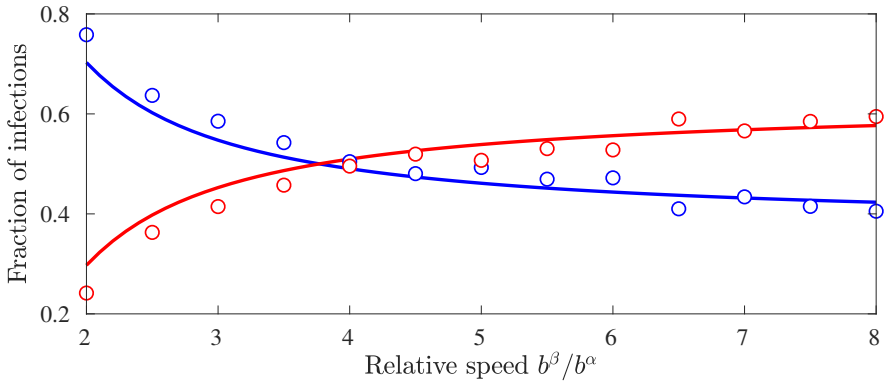


Figure 4: Fraction of nodes infected with strain α (blue) or β (red) in theory (lines) compared to simulations (circles) of an exclusive infection model on an Erdős-Rényi random graph with $N = 1000$ and $c = 3$. Other parameters are: $m^{\alpha\beta} = 0.01$ and $m^{\beta\alpha} = 0.25$, $d = 2$.

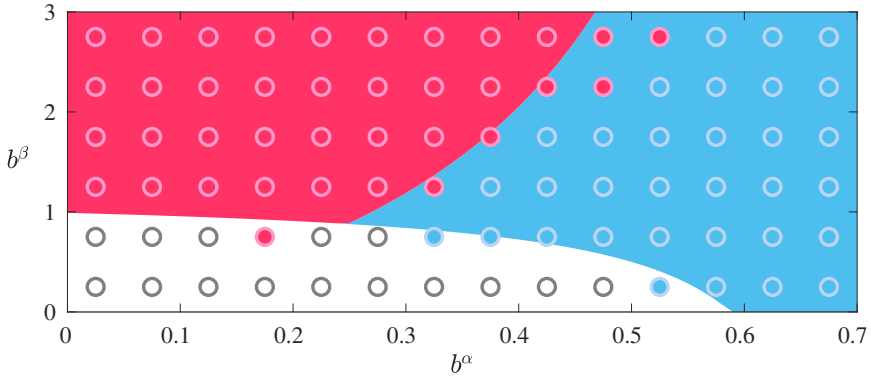


Figure 5: Phase diagram showing parameter regions in which there is no outbreak (white), or where an outbreak occurs and is dominated by either strain α (blue) or β (red). Coloured circles show the modal result of 100 simulated outbreaks on an Erdős-Rényi random graph with $N = 1000$ and $c = 3$. Other parameters are: $m^{\alpha\beta} = 0.01$ and $m^{\beta\alpha} = 0.33$, $d = 1$.

Several promising areas of development for our theory are open to investigation. We have restricted attention here to models in which (i) recovery times are independent between co-infecting strains, and (ii) both recovery and infection times are memoryless (i.e. exponential) random variables specified by a rate parameter. Extensions to our theory that relax both of these assumptions are possible. Message passing equations can also be formulated for models of bootstrap percolation [34]. These models correspond to contagions dynamics whereby the infection may only spread to individuals once they have reached a critical number of infected neighbours. This concept is relevant for instance in the modeling of opinion dynamics.

What is useful in the potential application of our approach to the models mentioned above, beyond the existing techniques, is the way in which it is centred on network topology. As we have observed such information is perhaps most succinctly captured by the non-backtracking matrix which we are able to employ to great effect. This is worthy of mention as further evidence for the effectiveness of a powerful tool, only recently introduced into the field of network science, but already proven highly effective in measuring relevant structural properties, such as percolation threshold [25] and community structure [35].

In the course of our approach we potentially also gain much more information about the dynamics of an outbreak than just the end state. For instance we might see how a node's infection probability changes in time. However, in order to assess such quantities directly it would be desirable to be able to solve systems like (2.4) in an iterative manner similar to that done for (2.5). Attempting this, one quickly encounters the problem of being able to understand time-lag between successive infections and how this depends on the network structure. Progress in this direction would be a valuable step towards providing greater insight into the spread of contagious processes on networks.

Authors' Contributions. SM undertook the mathematical analysis and drafted the manuscript. TR conceived of the study, supervised the analysis, and finalised the manuscript.

Funding. TR is supported by the Royal Society. SM is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/L015684/1.

Acknowledgements. The authors are grateful to Fakhteh Ghanbarnejad for stimulating preliminary discussions.

References

1. Council WG Dollar Sales Growth of Gluten-containing Ancient Grains in The United States in 2014 by Category. www.statista.com/statistics/509704/dollar-sales-growth-gluten-containing-ancient-grains-us-by-category/.
2. Council WG Dollar Sales Growth of Gluten-free Ancient Grains in The United States in 2014 by Category. www.statista.com/statistics/509695/dollar-sales-growth-gluten-free-ancient-grains-us-by-category/.
3. Mäkelä MJ, Puhakka T, Ruuskanen O, Leinonen M, Saikku P, Kimpimäki M, Blomqvist S, Hyypiä T, Arstila P. 1998 Viruses and bacteria in the etiology of the common cold. *Journal of clinical microbiology* **36**, 539–542.
4. Melegaro A, Choi YH, George R, Edmunds WJ, Miller E, Gay NJ. 2010 Dynamic models of pneumococcal carriage and the impact of the Heptavalent Pneumococcal Conjugate Vaccine on invasive pneumococcal disease. *BMC infectious diseases* **10**, 90.
5. Kermark M, Mckendrick A. 1927 Contributions to the mathematical theory of epidemics. Part I. *Proc. r. soc. a* **115**, 700–721.
6. Barrat A, Barthelemy M, Vespignani A. 2008 *Dynamical processes on complex networks*. Cambridge university press.
7. Li D, Fu B, Wang Y, Lu G, Berezin Y, Stanley HE, Havlin S. 2015 Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences* **112**, 669–672.
8. Wang P, González MC, Hidalgo CA, Barabási AL. 2009 Understanding the spreading patterns of mobile phone viruses. *Science* **324**, 1071–1076.
9. Cai W, Chen L, Ghanbarnejad F, Grassberger P. 2015 Avalanche outbreaks emerging in cooperative contagions. *Nature physics* **11**, 936.
10. Blythe RA, Allen RJ et al.. 2013 Parasites on parasites: Coupled fluctuations in stacked contact processes. *EPL (Europhysics Letters)* **101**, 50001.
11. Moore S, Mörters P, Rogers T. 2018 A re-entrant phase transition in the survival of secondary infections on networks. *Journal of Statistical Physics* pp. 1–14.
12. Neuhauser C. 1992 Ergodic theorems for the multitype contact process. *Probability Theory and Related Fields* **91**, 467–506.
13. Kiss IZ, Miller JC, Simon PL et al.. 2017 Mathematics of epidemics on networks. *Cham: Springer* **598**.
14. Rogers T. 2015 Assessing node risk and vulnerability in epidemics on networks. *EPL (Europhysics Letters)* **109**, 28005.
15. Hethcote HW. 2000 The mathematics of infectious diseases. *SIAM review* **42**, 599–653.
16. Keeling MJ, Eames KT. 2005 Networks and epidemic models. *Journal of the Royal Society Interface* **2**, 295–307.
17. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355.
18. Bansal S, Grenfell BT, Meyers LA. 2007 When individual behaviour matters: homogeneous and network models in epidemiology. *Journal of the Royal Society Interface* **4**, 879–891.

19. Pearl J. 1982 *Reverend Bayes on inference engines: A distributed hierarchical approach*. Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles.
20. Mezard M, Parisi G, Virasoro MA, Thouless DJ. 1988 Spin glass theory and beyond. *Physics Today* **41**, 109.
21. Karrer B, Newman ME. 2010 Message passing approach for general epidemic models. *Physical Review E* **82**, 016101.
22. Kühn R, Rogers T. 2017 Heterogeneous micro-structure of percolation in sparse networks. *EPL (Europhysics Letters)* **118**, 68003.
23. Moore S, Rogers T. 2020 Predicting the Speed of Epidemics Spreading in Networks. *Physical Review Letters* **124**, 068301.
24. Hamilton KE, Pryadko LP. 2014 Tight lower bound for percolation threshold on an infinite graph. *Physical review letters* **113**, 208701.
25. Karrer B, Newman ME, Zdeborová L. 2014 Percolation on sparse networks. *Physical review letters* **113**, 208702.
26. Hashimoto Ki. 1989 Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties* pp. 211–280. Elsevier.
27. Martin T, Zhang X, Newman ME. 2014 Localization and centrality in networks. *Physical review E* **90**, 052808.
28. Richardson M, Agrawal R, Domingos P. 2003 Trust management for the semantic web. In *International semantic Web conference* pp. 351–368. Springer.
29. Rozemberczki B, Davies R, Sarkar R, Sutton C. 2019 GEMSEC: Graph Embedding with Self Clustering. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2019* pp. 65–72. ACM.
30. Leskovec J, Lang KJ, Dasgupta A, Mahoney MW. 2009 Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* **6**, 29–123.
31. Suri S, Watts DJ. 2011 Cooperation and contagion in web-based, networked public goods experiments. *PloS one* **6**, e16836.
32. Gog JR, Grenfell BT. 2002 Dynamics and selection of many-strain pathogens. *Proceedings of the National Academy of Sciences* **99**, 17209–17214.
33. Gog J, Swinton J. 2002 A status-based approach to multiple strain dynamics. *Journal of mathematical biology* **44**, 169–184.
34. Shrestha M, Moore C. 2014 Message-passing approach for threshold models of behavior in networks. *Physical Review E* **89**, 022805.
35. Krzakala F, Moore C, Mossel E, Neeman J, Sly A, Zdeborová L, Zhang P. 2013 Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences* **110**, 20935–20940.